Detecting Changes in the AR Parameters of a Nonstationary ARMA Process

GEORGE V. MOUSTAKIDES and ALBERT BENVENISTE
Institut de Recherche en Informatique et Systèmes Aleatoires, Avenue du General Leclerc, Rennes 35042, France

(Accepted for publication July 15, 1985)

We present a method for detecting changes in the AR parameters of an ARMA process with arbitrarily time varying MA parameters. Assuming that a collection of observations and a set of nominal time invariant AR parameters are given, we test if the observations are generated by the nominal AR parameters or by a different set of time invariant AR parameters. The detection method is derived by using a local asymptotic approach and it is based on an estimation procedure which was shown to be consistent under nonstationarities.

KEY WORDS: Detection of changes, nonstationary ARMA processes.

I. INTRODUCTION

The problem of detecting changes in the spectral parameters of processes is frequently encountered in practice. Most methods (likelihood or innovation based techniques) solve this problem by using a complete description of the spectral parameters, (see for example [2, 8, 11]). Sometimes this description is not possible for all parameters, this is for instance the case when nonstationarities are present.

In such situations one is mainly interested in those parameters that can be described, while regarding the others as nuisance. A problem of this type is the problem of vibration monitoring. Here one is interested in detecting changes in the vibrating modes of a system,
II. PROBLEM STATEMENT

Let us consider the following system of difference equations

\[ x_{t+1} = F x_t + W_{t+1}, \quad x_0 = 0 \]  
(1)

\[ y_t - c^T x_t \]

where \( F \) is a real square matrix of dimension \( m \), \( c \) is a real vector of dimension \( m \) and \( \{W_t\} \) is a sequence of zero mean independent nonstationary vectors. With the superscript "T" we denote the transpose. The model in (1) is often used to model real systems. The vector \( x_t \) is called the state of the system, \( W_t \) is the input and the scalar \( y_t \) the output (observations). The process \( y_t \) is an ARMA process and if we write it in this form then we have

\[ y_t = \sum_{j=1}^{m} \alpha_j y_{t-j} - \sum_{j=0}^{m} \beta_j x_{t-j} + \epsilon_t + \sum_{j=1}^{m} \beta_j \epsilon_{t-j+1} \]  
(2)

where \( \{\epsilon_t\} \) is a standard i.i.d. sequence. The vector \( \alpha = [\alpha_1, \ldots, \alpha_m] \) is the vector of the AR parameters and the vector \( \beta = [\beta_0, \ldots, \beta_m] \) is the vector of the MA parameters. The vector \( \epsilon_t \) is related only to the matrix \( F \). Its components are the coefficients of the characteristic polynomial of this matrix. Since the roots of this polynomial are the
DETECTING PARAMETER CHANGES

spectral modes of the system. The overall spectral information is contained in the vector \( x \). The vector \( \beta \) is related to the input \( W \) thus it is time varying. Notice that with the model in (1) we can have a MA part in (2) of order at most \( m - 1 \).

Since any change in the spectral modes reflects into a change in the vector \( x \) we concentrate on this vector. Thus we are interested in detecting changes in the vector \( x \) in a nonsequential way, assuming that we have available the observation sequence \( \{ y_n \} \). We will assume that a nominal vector \( x = x_0 \) is known and that we do not know the vector \( x \) after the change. More specifically we will assume that a collection \( \{ y_1, \ldots, y_n \} \) is given and we would like to decide between the two hypotheses: \( H_0: x = x_0 \) and \( H_1: x = x_0 \). As we said in the introduction we will follow an asymptotic approach, thus we suppose that under change, \( x \) is of the form \( x = x_0 + \theta \sqrt{n} \), where \( \theta \) is an unknown direction of change. Although for our test we will never use the matrix \( F \), we will assume that \( x_0 \) corresponds to some nominal \( F_0 \) and \( x_0 + \theta \sqrt{n} \) to \( F_0 + \Theta \sqrt{n} \). Let us now consider the following vector:

\[
U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - Y_i t_0) Z_i
\]

(3)

where \( Y_i^T = [x_1, \ldots, x_{n-1}] \) and \( Z_i = [x_1, \ldots, x_{n-1}, u_{i+1}, \ldots, u_n] \). Equation (3) can be used to estimate the AR parameters of the observation process. Since \( U_n \) (disregarding \( 1/\sqrt{n} \)) is the sample correlation of the MA part with delayed enough observation and since this correlation is zero, we can estimate \( z \) by solving the system \( U_n = 0 \) (in the mean square sense if \( M > m \)). This estimate is known to be consistent even under nonstationarities [4]. We use a similar idea to derive a test. We expect that under no change the vector \( U_n \) will have a zero mean and under change a nonzero mean. Indeed we have

\[
\Sigma_m \rightarrow 2U_n \Leftrightarrow N(0, I) \quad \text{under } H_0
\]

\[
\Sigma_m \rightarrow (U_n - H_\alpha I) \Leftrightarrow N(0, I) \quad \text{under } H_1
\]

(4)

where \( \Delta \) means convergence in distribution, \( N(0, I) \) is the standard vector Gaussian density and \( \Sigma_m \) is the covariance matrix of \( U_n \). The
matrix $H_i$ is defined as

$$H_i = \sum_{k=1}^{\infty} \delta_k Y_k^I.$$  

(5)

We can see that under both hypotheses $U_i$ has asymptotically the same variance $\Sigma$, but not the same mean. Next step is to use $U_i$, as if we had equality in (4) and define a log-likelihood

$$T_i = U_i^T \Sigma^{-1} H_i \beta - 0.5 U_i H_i^T \Sigma^{-1} H_i \beta.$$  

(6)

Notice that $T_i$ is of length $M$ thus $H_i$ is of dimension $M \times \pi$. Assuming $M \geq 2 k$, let us substitute $\beta$ by its maximum likelihood estimate, this yields

$$T_i = U_i^T \Sigma^{-1} H_i \hat{H}_i \Sigma^{-1} H_i \beta - 0.5 U_i H_i^T \Sigma^{-1} H_i \beta.$$  

(7)

The quantity $T_i$ will be our test statistic. To define the threshold we have that under $H_0$, $T_i$ is asymptotically $\chi^2$ with $m$ degrees of freedom. Under $H_1$, $T_i$ is a centered $\chi^2$ with noncentricity equal to $0.5 H_i^T \Sigma^{-1} H_i$. Our test will be able to detect the change if the noncentricity is nonzero. As we will show in Section IV, with our assumptions this is always the case. The covariance matrix $\Sigma$ is not known and thus we must estimate it. We can for example use the sample covariance matrix defined as

$$\Sigma = \sum_{i=1}^{n} \sum_{j=1}^{n} (v_i - \bar{v})(v_j - \bar{v}) \Sigma_{ij},$$

(8)

where $v_i = y_i - \bar{y}_i$. Thus everything is defined in terms of known things. In the next sections we will prove the validity of (4), i.e. that the Central Limit Theorem (CLT) holds for $U_i$ under both hypotheses. Notice also that in (7) we take the inverse of the matrices $\Sigma$ and $H_i^T \Sigma^{-1} H_i$. We will show that these two matrices are full rank. For $\Sigma$, this will be necessary for the proof of the CLT, for the other matrix it is important in showing that the noncentricity factor is nonzero. Notice that the second matrix will be full rank if $H_i$ is full rank and $\Sigma$ is nonsingular and bounded from above. The proofs will be presented in several steps.
III. ASSUMPTIONS

Let us now introduce our assumptions. Let \( \{x_t\} \) be an ARMA process generated by a system of the form of (1.1) or (2.1). Let \( F_n, x_n \) be the nominal values of \( F \) and \( x \). We assume:

A1. The matrix \( F_n \) is full rank and has all its eigenvalues distinct and strictly inside the unit circle.

A2. The matrix \( h = [cF_n^0, \ldots, F_n^{n_1} - c] \) is of full rank.

A3. There exists a real \( p > 0 \) such that for every vector \( k \) and every integer \( n \) we have \( E[|k|^2 | x_n^k |^2] \leq n^{2p}, \) where by \( E[| \cdot |] \) we denote expectation.

A4. If \( Q_k = E[|W_k W_n^T|] \) is the input covariance matrix at time \( k \) and if \( Q_{\ast} = (1/n)^{r_0} Q_n \) is the average input covariance up to time \( n \), we assume that there exists a real \( \delta > 0 \) such that for every eigenvalue \( \gamma \) \( \lim_{n \to \infty} (\gamma - r_0)^{-1} (1 - F_n)^{-1} Q_{\ast} (1 - F_n)^{-1} (\gamma - r_0)^{-1} \geq \delta. \)

A5. There exists a nonzero vector \( z \) such that for every \( k \) we have \( Q_k \leq s^2 z z^T \).

Assumption A1 is to ensure the stability of our system, i.e. bounded inputs will produce bounded outputs. A2 is an observability condition, it ensures that any mode excited by the input will be observed in the output. Assumption A3 is rather technical; it requires that the input has uniformly bounded fourth-order moments. A4 means that in the average the input excites all the modes of the system and that these modes are present in the second order statistics. Finally Assumption A5 is to ensure that the input excites only at least at the direction of the vector \( z \).

Discussion. We will comment now on our assumptions. In A1 the assumption that the eigenvalues are distinct was made only to facilitate certain proofs; it can be relaxed to include multiple eigenvalues. Notice also that we assume knowledge of the exact system order \( m \). A2 seems necessary. Assumption A3 is quite strong, probably it can be relaxed to conditions involving only second order
moments. Assumption A4, as we said before, ensures that all the modes are present in the second order statistics. This assumption is very important since our test is based on second order statistics. If for example A4 was not true for some mode, in other words, if a mode even though excited by the input was not present in the second order statistics (this is the case when the MA part of a stationary ARMA process has zeros that are mirror images, with respect to the unit circle, of poles of the AR part), then any change on this mode cannot be detected by simply using second order statistics. Notice also that A4 involves only the average input covariance, that is, instantaneously we can have cancellation of modes. As one can see this assumption will be the base for proving that $H_0$ is of full rank. In the stationary case usually the assumptions up to A4 are sufficient to show the same things we like to show here. When nonstationarities are present this is no longer true, one can find examples where only with the first four assumptions we have a nonstationary covariance matrix $\Sigma$. Thus, it seems that an assumption of the form of $A3$ is also necessary.

IV. RESULTS

Before going to the proof of the CLT we will first present some lemmas that will be useful for this proof. Let us denote by $A$ and $B$ two real square matrices of dimension $m$ with distinct eigenvalues. Let $\lambda$ denote the eigenvalue of a square matrix $X$ with the maximum magnitude and $D$ a constant that depends only on $X$.

**Lemma 1** There exist constants $D_A, D_{A,B}$ such that for any two vectors $a, b$ and any integer $k \geq 0$ we have

\[ |A^k(AB)b| \leq D_A \lambda^k (\lambda(a^2)\lambda(b^2))^{1/2} \]

\[ |A(AB)^k - B(AB)^k| \leq D_{A,B} \lambda_{A,B} \lambda_{A,B} \lambda(a^2) \lambda(b^2) \]

where $\lambda(X)$ denote the maximum singular value of the matrix $X$.

**Proof** The proof of (i) is easy. Notice that it is obvious when $A$ is diagonal. When it is not diagonal we make a diagonalization and
Lemma 2. Let $F$, $W_i$ satisfy $A1$ and $A3$, then we also have that $X_i$ has uniformly bounded fourth order moments.

Proof. From (1) we have that

$$X_i = \sum_{k=0}^{K} F W_{k-1}$$

thus using the fact that $\{H_j\}$ is a sequence of independent vectors, we have for every vector $\lambda$ that

$$E(\langle X_i, \lambda \rangle^4)$$

$$= \sum_{j=0}^{K} \sum_{k=0}^{K} \sum_{l=0}^{K} \sum_{m=0}^{K} E(\langle F W_j, \lambda \rangle \langle F W_k, \lambda \rangle \langle F W_l, \lambda \rangle \langle F W_m, \lambda \rangle)$$

$$= \sum_{j=0}^{K} \sum_{k=0}^{K} \sum_{l=0}^{K} \sum_{m=0}^{K} E(\langle F W_j, \lambda \rangle^2) E(\langle F W_k, \lambda \rangle^2) E(\langle F W_l, \lambda \rangle^2)$$

$$\leq \sum_{j=0}^{K} \sum_{k=0}^{K} \sum_{l=0}^{K} \sum_{m=0}^{K} E(\langle F W_j, \lambda \rangle^2)^2.$$
The state covariance matrix of a stationary system with input covariance $Q$, assumed constant.

Proof: Since $W_t$ has uniformly bounded fourth-order moments, it will also have uniformly bounded second order moments, thus from (9), assuming $Q_0 = 0$ we have

$$P_t = \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} F^T P_{t-I} F^T = \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} F^T \left( \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} Q_{t-I} \right) F^T,$$

$$= \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} F^T Q_{t-I} F^T = \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} F^T Q F^T = \frac{1}{n^{1/2}} \sum_{I=1}^{n^{1/2}} Q_{t-I}. \tag{10}$$

The first term in (10) is equal to $F$, the second is uniformly of order $\|Q\|^2/n$ and the third term is of order $1/n$. Thus both last terms tend to zero. And this concludes the proof. Using this lemma we can show that we can approximate (order $1/n$) any average of matrices of the form $V_n = \frac{1}{n} \sum_{i=1}^{n} E[(Y_{t_i} \ldots Y_{t_j})^T (Y_{t_i} \ldots Y_{t_j})]$, by the corres-ponding matrix $V$ of a stationary system with input covariance $Q$, (assumed constant). This is true because $E(\{Y_{t_i} \ldots Y_{t_j}\}^T F) = 0$, and thus we can approximate every term in $V_n$ by the corresponding stationary term.

Lemma 4 Let $Q, Q'$ be two nonnegative definite matrices with $Q \geq Q'$. Let $F$ satisfy (4.1). Consider the following two systems in stationary situation.

$$X_{t+1} = FX_t + W_{t+1}, \quad y_t = g^T X_t$$

$$X_{t+1} = FX_t + W_{t+1}, \quad y_t = g^T X_t$$

with $W_t$ having covariance $Q$ and $W_t'$ having covariance $Q'$. For any integer $N$ we then have

$$E[(Y_{t_0} \ldots Y_{t_N})^T (Y_{t_0} \ldots Y_{t_N})] \geq E[(Y_{t_0} \ldots Y_{t_N})^T (Y_{t_0} \ldots Y_{t_N})]. \tag{12}$$

Proof: The proof is easy. Since we consider only second order statistics we can decompose $W_t$ into two independent processes $W_t, R_t$ such that $E(W_t R_t^T) = Q, E(R_t R_t^T) = Q' - Q$ and $E(W_t R_t^T) = 0$. Because of linearity, the process $y_t$ can be also decomposed into two
independent processes one due to $K_1$ and the other to $R_2$. The process due to $K_1$ will have exactly the same second order properties with $y_2$. Thus we conclude that the left hand side of (12) is equal to the right hand side plus another nonnegative term due to $R_2$.

**Lemma 5** Let $F, F'$ be two matrices satisfying $A1$ and $W_k$ a process satisfying $A2$. Let $X_k$ be the state vector of a system of the form of (1) and $X_k$ another state vector of a similar system but $F$ replaced by $F'$. For any vector $\lambda$ with $\lambda^T \lambda = 1$ we have

$$E[(X_k - \lambda^T X_k)^2] = 0 ||F - F'||^2,$$

uniformly in $k$. Where by $z = o(x)$ we mean $z$ is of the order of $x$.

**Proof.** By writing $X_k = \frac{1}{n} \sum_{i=1}^{n} F_i W_k, v_i$ using uniform boundedness of $\{Q_i\}$ and Lemma (6) we have

$$E[(X_k - \lambda^T X_k)^2] = \frac{1}{n} \sum_{i=1}^{n} E[(F_i - (F'))^T Q_i \lambda (F_i - (F'))^T \lambda]$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 E[(F_i - (F'))^T (F_i - (F'))]$$

$$\leq D ||F - F'||^2 \sum_{i=1}^{n} \lambda_i^2 = D ||F - F'||^2.$$

And this concludes the proof of the lemma.

**Lemma 6** Let $s$ be a nonzero vector, let $F$ satisfy $A1$ and $A2$ and let $E[W_k W_k^T] = \sigma^2 I$. Consider the system in (1) in the stationary situation, then for any fixed integer $N$ the covariance matrix

$$V = E[(\sum_{k=N+1}^{2N} [X_k - \lambda^T X_k])]$$

is of full rank.

**Proof.** Let $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ be any vector, then

$$\sum_{k=N+1}^{2N} \lambda_i^2 = \frac{1}{2\pi} \int_0^{2\pi} |\lambda_1 + \cdots + \lambda_n e^{i\omega t}|^2 |e^{-i\omega t} - F|^{-1} e^{i\omega t} dt.$$
If \( \lambda^j V_j = 0 \) then we will also have that the product under the integral is identically zero. Since the first term in this product is a trigonometric polynomial, if it is not identically zero, \( \sum \) \( a_n \) have only a finite number of zeros in the interval \((0,2\pi]\). Thus if we assume that \( \lambda^j \neq 0 \) we conclude that \( c_i(e^{-\lambda^j F_j} - \lambda^j)^{-1}s \) must be identically zero. But we have that
\[
c_i(e^{-\lambda^j F_j} - \lambda^j)^{-1}s = \sum_{i \in \mathbb{Z}} c_i(e^{-\lambda^j F_j} - \lambda^j)^{-1}s_i F_i.
\]
The above can be identified zero only if \( c_i F_i = 0 \) for every \( i \). This is not possible since because of \( \Delta^2 \) we have that for at least one \( i \) between zero and \( m-1 \) we have \( c_i F_i = 0 \), contradiction. Thus \( V \) is of full rank.

With the next theorem we prove the Law of Large Numbers for expressions that appear in the proof of the CLT.

**Theorem 1** Let \( F, W_k \) satisfy A1 and A3, let \( X_i \) be the state vector of a system of the form of (1). If \( [a_k], [b_k] \) are two uniformly bounded real vector sequences, then for any fixed integer \( j \)
\[
A_n = \sum_{i=1}^{n} b_i^k[X_{a+i}, X_{a+i}^T] - E[X_{a+i}, X_{a+i}^T]a_i \to 0 \quad a.s.
\]

**Proof** Since \( X_{i+a} = F_i X_i + \sum_{j=1}^{i-a} F_i^{-1}W_{j+a} \) and \( W_{j+a} \), is independent of \( X_i \), for \( i \geq 1 \) we have
\[
A_n = \sum_{i=1}^{n} \frac{1}{n} b_i^k [F_i^{j}, W_{j+a}, X_{a+i}]a_i + \sum_{i=1}^{n} \frac{1}{n} b_i^k [F_i^{j}, X_{a+i}, X_{a+i}^T]a_i - E[X_{a+i}, X_{a+i}^T]a_i.
\]
(13)

The sum of the form \( \sum_{i=1}^{n} \frac{1}{n} b_i^k [F_i^{j}, W_{j+a}, X_{a+i}]a_i \) are martingales. Following Feller ([5], page 245) in order to show that these normalized by \( n \) go a.s. to zero, it is enough to show that
\[
\frac{1}{n^2} \sum_{i=1}^{n} E[|b_i^k F_i^{j} W_{j+a}, X_{a+i}^T a_i|^2] < \infty.
\]
(14)

But (14) is true because
\[
E[|b_i^k F_i^{j} W_{j+a}, X_{a+i}^T a_i|^2] = E[|b_i^k F_i^{j} W_{j+a}|^2] E[|X_{a+i}^T a_i|^2]
\]
and from Lemma 2 we have that both terms in the product are uniformly bounded. Thus the first term in (13) go to zero a.s. To show that this is also true for the last term we will use the theory of martingales (see Hall [77], page 41). If \( \{X_n\}_{n=1}^\infty \) is a sequence of \( \sigma \)-algebras on \( \mathfrak{X} \), the \( \sigma \)-algebra generated by \{\( W_1, \ldots, W_l \)\} and if \( \mathcal{F}_n = \mathcal{F}^n [X_1, X_2^\perp - E(X_1 X_2^\perp)] \mathcal{F}^0 \), then if we can find \( \psi_n \rightarrow 0 \) and \( f_n \geq 0 \) such that

1. \( \| \psi_n - E[\psi_n \mathcal{F}_n] \|_2 \leq \psi_n + f_n \)
2. \( \| f_n \|_2 \psi_n \)
3. \( \sum_{n=1}^\infty f_n < \infty \) and \( \psi_n = o(k^{-\frac{1}{2} \log k})^{-2} \)

then \( \sum_{n=1}^\infty \psi_n = 0 \) a.s. For a proof of this statement see Hall ([77], page 41). We will show that we can define \( f_n \) and \( \psi_n \) to satisfy conditions 1 through 3. First notice that 1 is trivially satisfied for any nonnegative \( \psi_n, f_n \) because \( E[\psi_n \mathcal{F}_n] = \psi_n \). Let us now for simplicity denote \( d_n^2 = \mathcal{F}_n \), then \( d_n \) is also uniformly bounded. We can see that

\[
E[\psi_n \mathcal{F}_n] = d_n^2 (X_n X_n^\perp - E(X_n X_n^\perp)) (F_n)^n \psi_n
\]

Using the fact that the second moment of a random variable is larger than the variance we have from (15)

\[
\| \psi_n \mathcal{F}_n \|_2 \leq E[\| d_n^2 (X_n X_n^\perp - E(X_n X_n^\perp)) (F_n)^n \|_2] \]

\[
\leq E[\| d_n^2 (X_n X_n^\perp) \|_2] + E[\| (F_n)^n \|_2]
\]

Using now Lemma 1(0) and Lemma 2 we have from (16) that there exists a constant \( D \) such that \( \| \psi_n \mathcal{F}_n \|_2 \leq D \). We can thus define \( f_n = \psi_n \) and \( \psi_n = \psi_{a_n} \). Clearly with this definitions we have validity of conditions 2 and 3 and thus the Strong Law of Large Numbers holds also for the last term in (13).

We are now ready to prove the CLT defined by (4). This is done in the following theorem.

**Theorem 2** Let \( \{Y_i\} \) be the observation process defined by a system of the form of (1). If \( F_0 \) and \( W_0 \) satisfy conditions A1 through A5, then (4) holds.
Proof. We first show the C.L.T under hypothesis $H_0$. We will use a version of the martingale C.L.T, thus we will put $U_n$ defined in (3) under a martingale form. Let $k$ be a vector with $k^T x = 1$. What we would like to show is that $E[(X^T k^T k X)^2] = 1$. Notice that

$$E[(X^T k^T k X)^2] = 1,$$  

(17)


 this is true because $\Sigma_x$ was defined as the covariance of $U_n$. From (17) we see that we have the right variance required by the C.L.T. Dropping for simplicity the subscript "0" we have from (1)

$$\sum_{j=1}^{m-j} c_j F^{m-j-1} X_{t-k-j} + c_j \sum_{j=0}^{m-1} F^j W_{t-j}$$  

(18)

Substituting (18) in (3) and using the Cayley-Hamilton theorem we have

$$\sum_{j=1}^{m-j} c_j F^{m-j-1} X_{t-k-j} + c_j \sum_{j=0}^{m-1} F^j W_{t-j} = \lambda_j X_{t-k-j}$$  

(19)

where $\lambda_j = \lambda_j \Sigma_x^{1/2}$ and $c_j = c_j [F^{m-j-1} - z_1 F^{m-j-2} - \cdots - z_n F]$. Let us for the moment assume that the covariance $\Sigma_x$ is uniformly bounded away from zero for large enough $n$. We will prove this statement in Theorem 3. With this assumption $\lambda_j$ is uniformly bounded. Now rearranging the sum in (19) we have

$$\sum_{j=1}^{m-j} c_j F^{m-j-1} X_{t-k-j} + c_j \sum_{j=0}^{m-1} F^j W_{t-j} = \lambda_j X_{t-k-j}$$  

(20)

Therefore we can drop the constant $c_j$ from the sum. Defining

$$v_n = [Z_0 c_1^T + \cdots + Z_{m-n-k-n} c_n^T] W_{t-k-n}$$  

(21)

In order to show the C.L.T it is enough to show Lindeberg's condition for $v_n$ and that

$$\frac{\sum_{j=1}^{m-j} c_j^2}{n} = 1$$  

(22)
in probability (see Hall [7], page 52).  Ledebtberg's condition is the following, for any \(\epsilon > 0\) we want

\[
\sum_{\epsilon n} p_{\epsilon n} \left| I(I_{\epsilon n} > \epsilon n) \right| \rightarrow 0
\]

where with \(I(I)\) we denote the indicator function of the set \(A\). To show Lindberg's condition we have

\[
p_{\epsilon n} \left( I(I_{\epsilon n} > \epsilon n) \right) \leq \frac{E[e^{x_{\epsilon n}^2}]}{\epsilon n^2}
\]

Applying now the Schwartz inequality as (21) yields: \(E[e^{x_{\epsilon n}^2}] \leq \left( \sum_{\epsilon n} \frac{1}{\epsilon n} \right)^2 \left( \sum_{\epsilon n} x_{\epsilon n}^2 \right)^{1/2} \). And thus using Lemma 2 and the independence of \(W_{\epsilon n} \), with the \(Z_{\epsilon n} \), vectors, we can easily see that \(E[e^{x_{\epsilon n}^2}] \) is uniformly bounded by a constant \(D\). This yields

\[
\sum_{\epsilon n} p_{\epsilon n} \left( I(I_{\epsilon n} > \epsilon n) \right) \leq \frac{D}{\epsilon n^2} \rightarrow 0,
\]

and thus Lindberg's condition holds. To show (22) since using (17) and (20), we have \(1/\epsilon n \sum_{\epsilon n} E[x_{\epsilon n}^2] = \left( \sum_{\epsilon n} \frac{1}{\epsilon n} \right)^2 \left( \sum_{\epsilon n} x_{\epsilon n}^2 \right)^{1/2} + O(1/n) = 1+O(1/n) \), is enough to show that \(1/\epsilon n \sum_{\epsilon n} x_{\epsilon n}^2 \rightarrow 0 \) in probability. The random variable \(x_{\epsilon n}^2\) is a finite sum of terms of the form \(h \lambda_{\epsilon n} \lambda_{\epsilon n} \) where \(h\) depends on the vector \(\lambda_{\epsilon n}\).

Thus we can write

\[
\frac{1}{n} \sum_{\epsilon n} \left( \lambda_{\epsilon n} \lambda_{\epsilon n} \right) (e^{x_{\epsilon n}^2} + E[e^{x_{\epsilon n}^2}]) - E[\lambda_{\epsilon n} \lambda_{\epsilon n} E[e^{x_{\epsilon n}^2}]],
\]

\[
= \frac{1}{n} \sum_{\epsilon n} b_{\epsilon n} \lambda_{\epsilon n} \lambda_{\epsilon n} \left( e^{x_{\epsilon n}^2} - E[e^{x_{\epsilon n}^2}] \right),
\]

\[
= \frac{1}{n} \sum_{\epsilon n} b_{\epsilon n} \left( e^{x_{\epsilon n}^2} - E[e^{x_{\epsilon n}^2}] \right) \lambda_{\epsilon n} \lambda_{\epsilon n} - \lambda_{\epsilon n} \lambda_{\epsilon n} - \left( \lambda_{\epsilon n} \lambda_{\epsilon n} \right) E\left( e^{x_{\epsilon n}^2} \right).
\]

The first sum in (23) is a martingale. Using similar reasoning as in Theorems 1 we can show that it converges to zero a/s. For the second term in (23) we apply Theorem 1 and thus this term also goes to
zero a.s. With these arguments we have shown that (22) is true not only in probability but a.s.

Up to this point we have shown the CLT under $H_0$. To prove it also under $H_1$, notice that the observation sequence $\{X_i\}$ is now generated not by the nominal matrix $F_0$ but by the matrix $F$ that satisfies $\|F-F_0\|=\sqrt{1/\alpha}$. Let us call $\{Y_i\}$ the observation sequence generated by the nominal matrix $F_0$ when as input we have exactly the same like the system with $F$. If we show for every vector $x$ with $x^T Z = 1$ that we have

$$
E\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \bar{Y}) Y_i Z_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \bar{Y}) Y_i Z_i^T\right]^T \to 0,
$$

(24)

where $Y_i$, $Z_i$ are the $Y_i$, $Z_i$ vectors corresponding to the nominal system, this means that the two terms in (24) have the same asymptotic distribution. Since the second term, as it was shown in the first part of this theorem, is asymptotically Gaussian, the same will be true for the first term as well. To show (24) define

$$
a_k = (y_{k+1} - x^T Y_k) e^T Z_k - (y_{k+1} - x^T Y_k) e^T Z_k^T.
$$

The terms $y_{k+1} - x^T Y_k$ and $y_{k+1} - x^T Y_k$ are the MA parts of the two observation processes and since they are at most $(m-1)$-dependent we have $E(a_k a_j^T) = 0$ for $j \neq m$. Thus (24) is equivalent to

$$
\frac{1}{n} \left( \sum_{k=1}^{n} a_k^T \right) = \frac{1}{n} \sum_{k=1}^{n} E(a_k^T) + \frac{1}{n} \sum_{j=1}^{n-1} \sum_{k=1}^{n-j} E(a_k a_j^T) \to 0. \quad (25)
$$

Since $2|E(a_k a_j^T)| \leq E(a_k a_j^T) + a_k^T a_j$, in order to prove (25) it is enough to prove that the first term in (25) goes to zero. Notice that we have

$$
a_k = (y_{k+1} - x^T Y_k) [E(Z_k Z_k^T)] + (y_{k+1} - x^T Y_k) (y_{k+1} - x^T Y_k) [E(Z_k Z_k^T)]
$$

The two MA parts are independent of $Z_k$ and $Z_k^T$, thus

$$
E(a_k^T) \leq 2E[(y_{k+1} - x^T Y_k) E(Z_k Z_k^T)]^T + 2E[(y_{k+1} - x^T Y_k) E(Z_k Z_k^T)]^T.
$$

(26)
Using Lemma 5 the quantities $y_{i, 1}, y_1$ and $Z_i$ differ from the corresponding normal quantities (in the mean square sense) by an amount $0(<F_{0, 0}^{1/2}) - (0.1/n)$. Thus for every $k < n$ we have $E[|y_{i, 1}|] = O(1/n)$ which yields $\sum_{i=1}^{n} E[|y_{i, 1}|] = O(1/n)$. And thus we have shown that (24) is true. What is left now to prove is in order for Theorem 2 to be complete is that the covariance matrix $\Sigma$ is uniformly bounded away from zero. This is shown in Theorem 3.

**Theorem 3** Let $E_0, \Sigma_0$ satisfy A1 through A5. Let $\Sigma_0$ be the covariance matrix of $U_{0,i}$ defined in (3) and $H_{i,i}$ the matrix defined in (5), then for large enough $i$ the matrix $\Sigma_i$ is uniformly full rank and uniformly bounded from above also $H_i$ is a.s. uniformly full rank.

**Proof** The proof is based on certain properties that hold under stationarity. From Eq. (20) we have that

$$U_{i, i} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left[ Z_k e_{i, 1}^k + \cdots + Z_{k-m} e_{i, m}^k \right] W_{i, k-i} + \left( \frac{1}{\sqrt{n}} \right).$$

We first prove that $\Sigma_i$ is full rank. Let $\lambda$ be a vector with $\lambda^T \Phi = 1$. Let us call $t_0 = \Phi^T Z_0$ and $T_i = [t_{i, 1}, \ldots, t_{i, n-m}]$, then

$$\lambda^T U_{i, i} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} T_i^T W_{i, k-i} + \left( \frac{1}{\sqrt{n}} \right),$$

where $T_i = [t_{i, 1}, \ldots, t_{i, n-m}]$. Using A5 and that $W_{i, i+1}$ is independent of $T_0$, we have

$$E[T_i^T U_{i, i}^2] = \frac{1}{n} \sum_{k=1}^{n} E[T_i^T T_0 Z_k^2] + \left( \frac{1}{n} \right)\sum_{k=1}^{n} E[T_i^T T_0^2] + \left( \frac{1}{n} \right).$$

(27)

and let $t_0 = 0$. Note that $E$ comes from the matrix $0$ defined in A2 by using linear operators. Since $0$ is assumed full rank so is $E$, thus the vector $t_0$ is nonzero. Let $E = [t_{0, 1} \ldots t_{0, n-m}]$ and $\lambda^T = [\lambda_{0, 1} \ldots \lambda_{0, n-m}]$, then we can see that

$$T_i^T = [\lambda_{0, 1} \ldots \lambda_{0, n-m}]$$

(28)

$$\begin{bmatrix}
\lambda_{0, 1} & 0 & 0 & \cdots & 0 \\
0 & \lambda_{0, 2} & 0 & \cdots & 0 \\
0 & 0 & \lambda_{0, 3} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \lambda_{0, n-m}
\end{bmatrix}$$

}\]
Denote with \( L \) the matrix in (28). This matrix \( L \) is of dimensions \( M \times (M + m - 1) \). It is easy to see that \( L^T L \) gives the convolution of the two sequences \( \{a_1, \ldots, a_N\} \) and \( \{b_1, \ldots, b_M\} \). It is known that the convolution of two sequences cannot be identically zero unless one of the two sequences is. Since \( \sigma_L > 0 \) means that there is no \( L^T L = 0 \) such that \( L^T L = 0 \) or, that \( L \) is of full rank. Let us call \( \sigma_L > 0 \) the smallest singular value of \( L^T L \).

Now define

\[
P_L = \frac{1}{n} \sum_{n} E[\{y_1 \ldots y_{n+m-L+1}\}^T \{y_1 \ldots y_{n+m-L+1}\}],
\]

(29)

the average covariance matrix of the random vector in (28). From Lemma 3 \( P_L \) can be approximated (order 1/\( n \)) by the corresponding covariance of a stationary system with input \( C_0 \). Since from A5 we also have \( C_0 \approx \sigma_L^2 \), using Lemma 4 this last covariance can be lower bounded by the corresponding covariance \( V \) of a stationary system with input \( \sigma_L^2 \). The covariance \( V \) from lemma 6 is of full rank. Call \( \sigma_V > 0 \) the smallest singular value of \( V \). Going back to (29) and using (28) and (29) we have

\[
E[\{X^T U_a \}^2] \geq \lambda_1^2 E[\{X_a \}^2] / \sigma_L^2
\]

(30)

and thus for large enough \( n \), \( \Sigma_n \) is uniformly bounded away from zero. Notice that for the proof of the nonsingularity of \( \Sigma_n \) we did not use A4. To prove that \( \Sigma_n \) uniformly bounded from above we proceed in a similar way but we use the uniform bounds of the covariance matrices involved. We omit the details.

To prove that \( H_a \) is a.s. of full rank notice that because of Theorem 1 we have that \( H_a - E[\hat{H}_a] \to 0 \) a.s., thus it is enough to show that \( E[\hat{H}_a] \) is of full rank. Using Lemma 4, \( E[\hat{H}_a] \) can be approximated by the corresponding matrix \( H \) of a stationary system with input \( C_0 \). We will show that this matrix \( H \) is of full rank. We have that \( H = E[\{Y_i \}^2 \] \), where \( Y_i \) and \( Z_i \) are now generated by a stationary system with input \( C_0 \) (assumed constant). We would like to show that there exists real \( a > 0 \) such that for any vector \( \lambda \) with \( \lambda_x \neq 1 \) we have \( \lambda^T H \lambda \geq a \). Since \( Z_i \) has length \( M \geq m \) we consider only the first \( m \) components of this vector, that is, we will show that \( E[\{Y_i \}^2 \] \) is of full rank. Consider the vector \( E[\{X^T Y_{-1} \}^2] \),
using similar ideas as in [1, 19], the $i$th component of this vector (say $d_i$) is given by
\[ d_i = \frac{1}{2\pi i} \int (z^{i-1} \cdot Q_d(z^{-1}I - F_d)^{-1}Q_d(z^{-1}I - F_d))^{-1} e^{	au z^{-1}i} \frac{dz}{z} \]
\[ i = 1, \ldots, m \] (31)
where $\hat{z}(z) = n z^{i-1} \cdot l_1 \cdot l_m$ and the integration path is the unit circle. The only poles inside the unit circle in (31) are the eigenvalues $r_i$ of $F_d$. Thus from (31)
\[ d_i = \frac{\hat{z}(r_i)}{r_i - r_i} \mu_i \rho_i \] (32)

where $\mu_i = \lim_{z \to r_i} (z - r_i)^{-1} \cdot Q_d(z^{-1}I - F_d)^{-1}Q_d(z^{-1}I - F_d)^{-1} e^{	au z^{-1}i}$. Since the Vandermonde matrix
\[ V_n = \begin{bmatrix} 1 & r_1 & \cdots & r_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_m & \cdots & r_m^{n-1} \end{bmatrix} \]
is of full rank, call $\sigma_i$ its smallest singular value. Notice that the vector $(\hat{z}(r_1), \ldots, \hat{z}(r_m))$ can be written as $T V_n$ where $T = [\tau \cdots \tau]$. Let us call
\[ \rho_i = \left[ \frac{\hat{z}(r_i)}{r_i - r_i} \mu_i \frac{\hat{z}(r_m)}{r_m - r_m} \mu_m \right]. \]
then using A4, i.e. that $|\mu| \geq \delta$ and also from A1 that $|\nu_i| < 1$ we have
\[ \sum_{i=1}^n d_i^2 = \rho_i^2 \sum_{i=1}^n |\hat{z}(r_i)|^2 \geq \delta^2 |\mu|^2 \rho_i^2 \sum_{i=1}^n |\hat{z}(r_i)|^2 \]
\[ = \tau^2 \sum_{i=1}^n |\hat{z}(r_i)|^2 \sum_{i=1}^n \frac{\hat{z}(r_i)}{r_i - r_i} \mu_i \rho_i \]
where the superscript $''\ast''$ we mean complex conjugate. And this
concludes the proof. As we have proved, \( \Sigma \) is uniformly bounded from above and \( H \) is uniformly of full rank, thus we have that \( H\Sigma^{-1}H^T \) is also uniformly bounded away from zero.

V. CONCLUSION

We have presented a method for detecting changes in the AR parameters of a nonstationary ARMA process. The detection scheme was derived by using the same idea that is used for the estimation of these parameters, i.e., that the MA part of an ARMA process is independent from delayed enough observations. Following a local asymptotic approach, the detection of a change in the AR parameters was reduced to the detection of a nonzero mean of a Gaussian random vector. With the assumptions we have introduced here we can actually show a stronger result, namely that the CLT in (4) remains valid if \( \Sigma \) is replaced by \( \Sigma_0 \). This is true basically because it is possible to show that the difference between the two matrices goes to zero u.s. The proof of this last statement is unfortunately long, thus we have decided to present only the weaker version of the CLT defined in (4). To say a few things about estimating the covariance matrix \( \Sigma \); even though \( \Sigma_0 \) has expectation equal to \( \Sigma \), it has the drawback that it is not always positive definite. This can lead to a negative test statistics \( \zeta \) (Eq. 7). For practical purposes we can only use the first term of \( \Sigma \), which is positive definite. This method was successfully used in detecting changes in vibrating modes of linear systems [3]. Finally another practical problem is the knowledge of the system order \( m \). In simulations our method performed very well even when the true order was overestimated.

References


